

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:41 ; Search time 170.72 Seconds

(without alignments)  
23.864 Million cell updates/sec

Title: US-09-331-631A-7\_COPY\_81\_140

Perfect score: 342  
Sequence: 1 LOROYQCCGRCOEQOQGOR.....HENYHNHKNRSEEEGQOR 60

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR66:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	342	100.0	566	2 S22477	vicillin precursor
2	133	38.9	509	2 S08059	alpha-globulin typ
3	133	38.9	505	2 S06398	alpha-globulin typ
4	132	38.6	588	1 FMCNAB	alpha-globulin B p
5	104	30.4	810	2 T44430	protein PVI00 (imp
6	103.5	30.3	1038	2 T02634	rep protein homolo
7	100	29.2	613	2 S27770	hypothetical prote
8	93	27.2	708	2 A53185	G-box-binding fac
9	93	27.2	930	2 T08588	hypothetical prote
10	92	26.9	139	2 A26892	Mopa box protein
11	91	26.6	1457	2 T14577	protein kinase yak
12	90.5	26.5	285	1 T46207	involucrin - dog
13	90.5	26.5	395	2 S35565	sex-determining pr
14	90.5	26.5	395	2 S43344	sex-determining pr
15	90	26.3	905	1 RGHV55	regulatory protein
16	90	26.3	1094	2 S49313	protein kinase - s
17	89	26.0	646	2 D82493	conserved hypotet
18	89	26.0	853	2 T46347	hypothetical prote
19	89	26.0	1010	2 T13167	Lotia-like protein
20	88	25.7	2150	2 S71629	sensory transducti
21	87	25.4	551	2 S51941	protein 1 precursor
22	87	25.4	1154	2 S69206	regulator protein
23	86	25.1	255	2 A60637	mezozoite antigen
24	86	25.1	411	2 T29475	legumin B - fava b
25	86	25.1	564	2 S37241	unconventional myo
26	85.5	25.0	1737	2 A59235	A-kinase anchor pr
27	85.5	25.0	2359	2 T03094	homeotic protein p
28	85	24.9	1403	2 S24548	gene mastermind pr
29	85	24.9	1655	2 T13998	

30	84.5	24.7	648	1 J01150	protein kinase (EC
31	84	24.6	1761	2 T13675	hypothetical prote
32	83.5	24.4	544	1 I36911	involucrin L - dou
33	83.5	24.4	905	2 T23229	hypothetical prote
34	83.5	24.4	919	2 A39248	androgen receptor
35	83	24.3	1090	2 A41696	regulatory protein
36	83	24.3	1180	2 S69205	stripe a/b protein
37	82.5	24.1	467	1 A49377	involucrin - mouse
38	82	24.0	1390	2 T14004	trfa protein - sil
39	82	24.0	1905	2 T18267	multidrug resistan
40	81.5	23.8	4857	2 T03455	ALR protein - huma
41	81.5	23.8	5262	2 T03454	ALR protein - huma
42	81	23.7	339	1 TWHU2D	transcription init
43	81	23.7	582	2 B53234	vicillin-like stora
44	81	23.7	758	2 S54522	hypothetical prote
45	81	23.7	816	2 S05548	gap protein hunchb

## ALIGNMENTS

```

RESULT 1
S22477
vicillin precursor - cacao
C:Species: Theobroma cacao (cacao)
C:Date: 30-Sep-1993 #sequence-revision 30-Sep-1993 #text-change 17-Mar-1999
C:Accession: S22477; S22478; S18105; S22050
R:McHenry, L., Fritz, P.J.
Plant Mol. Biol. 18, 1173-1176, 1992
A:Title: Comparison of the structure and nucleotide sequence of vicillin genes of coco
A:Reference number: S22477; MUID:92288309
A:Accession: S22477
A:Molecule type: DNA
A:Residues: 1-566 <MC2>
A:Cross-references: EMBL:X62625
A:Accession: S22478
A:Molecule type: mRNA
A:Residues: 1-452 <MC2>
A:Cross-references: EMBL:X62626
C:Genetics:
A:Introns: 211/1; 269/3; 296/3; 391/3; 502/1
C:Superfamily: glycinn
C:Keywords: seed; storage protein
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-566/Product: vicillin #status predicted <MAT>

Query Match 100.0%; Score 342; DB 2; Length 566;
Best Local Similarity 100.0%; Pred. No. 1.3e-25;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LOROYQCCGRCOEQOQGORCOQKCEQYKDEGRGHEHYHNHKNRSEEEGQOR 60
DB 81 LOROYQCCGRCOEQOQGORCOQKCEQYKDEGRGHEHYHNHKNRSEEEGQOR 140

RESULT 2
S08059
alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)
M:Alternate names: seed storage protein
C:Species: Gossypium hirsutum (upland cotton)
C:Date: 31-Mar-1990 #sequence-revision 31-Mar-1990 #text-change 30-Sep-1993
C:Accession: S08059
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX.
A:Reference number: S06398
A:Accession: S08059
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-509 <CHI>
C:Superfamily: glycinn

```

```
Query Match          38.9%; Score 133; DB 2; Length 509;
Best Local Similarity 35.8%; Pred. No. 1e-05;
Matches 29; Conservative 12; Mismatches 18; Indels 22; Gaps 3;

OY 2 QROYOQCGRCOEQOOGQREOQOCCORCKWEQYKEQ-----ERGEHENYNN 46
      |||:|:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 40 QROFEOCQOCHQOQOERPRKQOCVRECKREKYQENPWRGEREEAEETEGEOQOSH 99
      |||:|:| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 47 ----HK---NRSEEEGQOR 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 100 PPHFRRSFQSRFRHEGNFR 120.

RESULT 3
s06398
alpha-globulin type A precursor - upland cotton
N:Alternate names: seed storage protein
C:Species: Gossypium hirsutum (upland cotton)
C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
C:Accession: S06398
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Se
A:Reference number: S06398
A:Accession: S06398
A:Molecule type: mRNA
A:Residues: 1-588 <CHL>
A:Cross-references: GB:M1691; NID:g167374; PIDN:AAA33071.1; PID:g167375
A:Experimental source: var. Coker 201
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 7, 475-489, 1986
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination XVII c
A:Reference number: A30838
A:Accession: A30838
A:Molecule type: mRNA
A:Residues: 1-588 <CHL>
A:Cross-references: GB:M1691; NID:g167374; PIDN:AAA33071.1; PID:g167375
A:Experimental source: var. Coker 201
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Se
A:Reference number: S06398
A:Accession: S06398
A:Molecule type: DNA
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-81 <CH2>
C:Comment: This is a seed storage protein.
C:Superfamily: glycyltin
C:Keywords: glycoprotein; seed; storage protein
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-588/Product: alpha-globulin storage proprotein #status predicted <MAT>
F:417/Binding site: carbohydrate (asn) (covalent) #status predicted
```

```
Query Match          38.6%; Score 132; DB 1; Length 588;
Best Local Similarity 35.8%; Pred. No. 1.4e-05;
Matches 29; Conservative 11; Mismatches 19; Indels 22; Gaps 3;

OY 2 QROYOQCGRCOEQOOGQREOQOCCORCKWEQYKEQ-----ERGEHENYNN 46
      |||:|:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 QROFEOCQOCHQOQOERPRKQOCVRECKREKYQENPWRGEREEAEETEGEOQOSH 180
      |||:|:| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 47 ----HK---NRSEEEGQOR 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 PPHFRRSFQSRFRHEGNFR 201
```

```
RESULT 5
T44430
protein pV100 (imported) - winter squash
C:Species: Cucurbita maxima (winter squash)
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T44430
R:Yamada, K.; Shimada, T.; Kondo, M.; Nishimura, M.; Hara-Nishimura, I.
J. Biol. Chem. 274, 2563-2570, 1999
A:Title: Multiple functional proteins are produced by cleaving Asn-Gln bonds of a sin
A:Reference number: T22767; MUID:99107919
A:Accession: T44430
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-810 <YAM>
A:Cross-references: EMBL:AB019195; NID:g3808061; PIDN:BA434056.1; PID:g3808062
```

```
Query Match          30.4%; Score 104; DB 2; Length 810;
Best Local Similarity 31.6%; Pred. No. 0.0088;
Matches 18; Conservative 18; Mismatches 21; Indels 0; Gaps 0;

OY 4 QYQOQCGRCOEQOOGQREOQOCCORCKWEQYKEQERGEHENYNNHKKNRSEEEGQOR 60
      |||:|:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 77 EYEVCLRCQVAREGVEDQORCEQVCEERLREREDGVEDVERDPEWEREROR 133
```

```
RESULT 6
T02634
rep protein homolog - slime mold (Dictyostellium discoideum) plasmid Ddp5
C:Species: Dictyostellium discoideum
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 29-Oct-1999
C:Accession: T02634
R:Riepen Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; H
Genetics 146, 1117-1125, 1996
A:Title: Dictyostellium discoideum nuclear plasmid Ddp5 is a chimera related to the Dd
A:Reference number: Z14684; MUID:98198636
A:Accession: T02634
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1038 <RIE>
A:Cross-references: EMBL:AF000580; NID:g3068582; PIDN:AAJ14390.1; PID:g3068583
A:Experimental source: strain WS2162
C:Genetics:
A:Gene: rep
A:Genome: plasmid
A:Mobile element: plasmid Ddp5
```

```
Query Match          30.3%; Score 103.5; DB 2; Length 1038;
Best Local Similarity 37.3%; Pred. No. 0.012;
Matches 22; Conservative 17; Mismatches 19; Indels 1; Gaps 1;

OY 2 QROYOQCGRCOEQOOGQREOQOCCORCKWEQYKEQERGEHENYNNHKKNRSEEEGQOR 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 940 QROEQERERREOQOEOEQEOQOE-QOEOEQOEOQOEOEQOQOQOQOQOQOQOQOQOQOQO 997
```

RESULT 7

```

527770
hypothetical protein 1 - African malaria mosquito (fragment)
C:Species: Anopheles gambiae (African malaria mosquito)
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997
C:Accession: S27770
R:Basansky, N.J.; Paskewitz, S.M.; Mills-Hamm, D.M.; Collins, F.H.
submitted to the EMBL Data Library, June 1992
A:Description: Distinct families of site-specific retroposons occupy identical positions
A:Reference number: S27770
A:Accession: S27770
A:Molecule type: DNA
A:Residues: 1-613 <BES>
A:Cross-references: EMBL:M93690; NID:g159615; PID:g159616

Query Match          29.2%, Score 100; DB 2; Length 613;
Best Local Similarity 31.3%; Pred. No. 0.017;
Matches 21; Conservative 18; Mismatches 20; Indels 8; Gaps 1;

Oy      2   ORGYQCCGRCOEQQGGREQQCCORKCWE-----QYKEQERGEHNNHNKKNRSE 53
           :| | | | :||| |::| | | | | | | | | | | | | | | | | | | | | | | |
Db      232 REQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQVQQQQQQHQKQQ 291

Oy      54   EEEGQR 60
           :::: |||

Db      292 QQQQQQR 298

RESULT      8
A53185
G-box-binding factor - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-Sep-1999
C:Accession: A53185
R:Schnitzler, G.R.; Fischer, W.H.; Firtel, R.A.
Genes Dev. 8, 502-514, 1994
A>Title: Cloning and characterization of the G-box binding factor, an essential component
A:Reference number: A53185; MUID:94I70994
A:Accession: A53185
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-708 <SCH>
A:Cross-references: GB:I29075; NID:g456561; PID:AAA21021.1; PID:g456562
C:Superfamily: G-box binding factor
C:Keywords: DNA binding; transcription factor; zinc finger

Query Match          27.2%, Score 93; DB 2; Length 708;
Best Local Similarity 31.0%; Pred. No. 0.087;
Matches 18; Conservative 14; Mismatches 26; Indels 0; Gaps 0;

Oy      2   ORYYQCCGRCOEQQGGREQQCCORKCWQYKKEERESENHYNNKKNRSEERQQQ 59
           || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      190 QQHQQQQQQQQQQHHQQQQQQQQHHQQQQQQQQHHQQQQQQQQHHNHSPPQQQHNNQQQHNNQQQ 247

RESULT      9
T08588
hypothetical protein L23H3.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C:Accession: T08588
R:Bavan, M.; Pohl, T.; Weizenecker, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schue
submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16098
A:Accession: T08588
A:Molecule type: DNA
A:Residues: 1-930 <BEV>
A:Cross-references: EMBL:AL050398; GSPDB:GN00062; ATSP:L23H3.30
A:Experimental source: cultivar Columbia; BAC clone L23H3
C:Genetics:
A:Gene: ATSP:L23H3.30
A:Map position: 4

```

```

Oy: Introns: 11/2; 51/1; 87/3; 249/3; 278/2; 304/3; 330/1; 346/3; 449/3; 523/3; 605/3;

Query Match          27.2%; Score 93; DB 2; Length 930;
Best Local Similarity 27.9%; Pred. No. 0.11;
Matches 17; Conservative % 24; Mismatches 18; Indels 2; Gaps 1;

Oy      2 QROYQCCGCR--CQEQQQGREGQQCCQRKCKEYKDEQREGEHENTHNHKRNSSEEGCG 59
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      95 EQLDLSQHHPVSDQQDDQDDQDDQDDQDDQDDQDDQDDQDDQDDQDDQDDQDDQDDQ 154
Oy      Oy      60 R 60
       :
Db      155 Q 155

RESULT 10
A26892
Mopa box protein - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 05-Nov-1999
C:Accession: A26892
R:Duboule, D.; Haenlin, M.; Galliot, B.; Mohler, F.
Mol. Cell. Biol. 7, 2003-2006, 1987
A:Title: DNA sequences homologous to the Drosophila opa repeat are present in murine
A:Reference number: A26892; MUID:87257908
A:Molecule type: mRNA
A:Residues: 1-139 <DDB>
A:Cross-references: GB:M16362; NID:g200142; PIDN:AAA39860.1; PID:g387503

Query Match          26.9%; Score 92; DB 2; Length 139;
Best Local Similarity 30.5%; Pred. No. 0.027;
Matches 18; Conservative % 21; Mismatches 20; Indels 0; Gaps 0;

Oy      2 QROYQCCGRCQEQEQGGREGQQCCQRKCKEYKDEQREGEHENTHNHKRNSSEEGCG 60
       |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      7   QQNQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 65

RESULT 11
T14577
protein kinase Yaka (EC 2.7.1.-) - slime mold (Dictyostellum discoideum)
C:Species: Dictyostellium discoideum
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14577
R:Kuspa, A.; Lu, S.; Souza, G.M.
Submitted to the EMBL Data Library, January 1998
A:Description: Yaka, a protein kinase required for the growth to development transits
A:Reference number: Z18146
A:Accession: T14577
A>Status: preliminary; translated from GH/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1457 <KUS>
A:Cross-references: EMBL:AF045453; NID:g2854116; PID:g2854117; PIDN:AAC02554.1
C:Genetics:
A:Gene: yaka
A:Keywords: ATP; phosphoprotein; phosphotransferase; serine/threonine-specific protei
```





